



Prédiction du Churn Client - Projet 3 AIA

Rabearivony Tsanta Loïc

Formation : AIA01

Septembre 2025



Objectifs du projet

- Comprendre les **facteurs** qui influencent le churn
- Construire et comparer **plusieurs modèles de classification**
- Optimiser les **hyperparamètres** (GridSearchCV)
- Fournir des **améliorations business** exploitables



Données utilisées

- Source : **Kaggle – Telco Customer Churn**
- Nombre de clients : **~7 000**
- Variables : caractéristiques clients, services souscrits, facturation, churn (Oui/Non)
- Données nettoyées, encodées et normalisées

Analyse exploratoire (EDA)

Vue d'ensemble des Données

Nombre total de lignes

7032

Nombre total de colonnes

24

Utilisation mémoire

1.29 MB

Échantillon du jeu de données

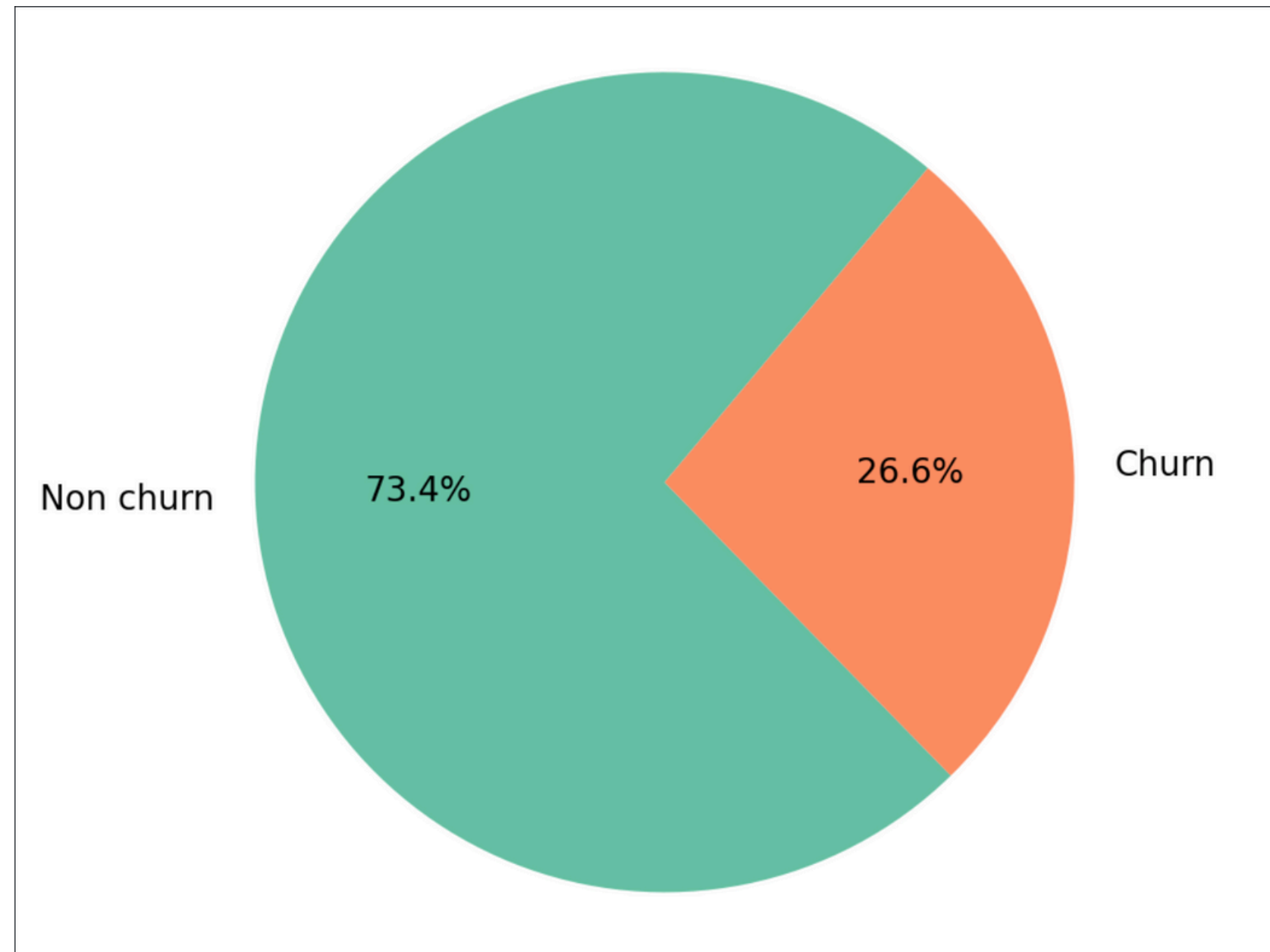
	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	PaperlessBilling	MonthlyCharges	TotalCh
0	0	0	1	0	-1.2802	0	0	0	1	0	0	0	0	1	-1.1617	-0
1	1	0	0	0	0.0643	1	0	1	0	1	0	0	0	0	-0.2609	-0
2	1	0	0	0	-1.2395	1	0	1	1	0	0	0	0	1	-0.3639	-0
3	1	0	0	0	0.5125	0	0	1	0	1	1	0	0	0	-0.7479	-0
4	0	0	0	0	-1.2395	1	0	0	0	0	0	0	0	1	0.1962	-0

Informations sur les colonnes

	Colonne	Type de données	Nombre de valeurs non nulles	Nombre de valeurs nulles
0	gender	int64	7032	0
1	SeniorCitizen	int64	7032	0
2	Partner	int64	7032	0
3	Dependents	int64	7032	0
4	tenure	float64	7032	0
5	PhoneService	int64	7032	0
6	MultipleLines	int64	7032	0
7	OnlineSecurity	int64	7032	0
8	OnlineBackup	int64	7032	0
9	DeviceProtection	int64	7032	0

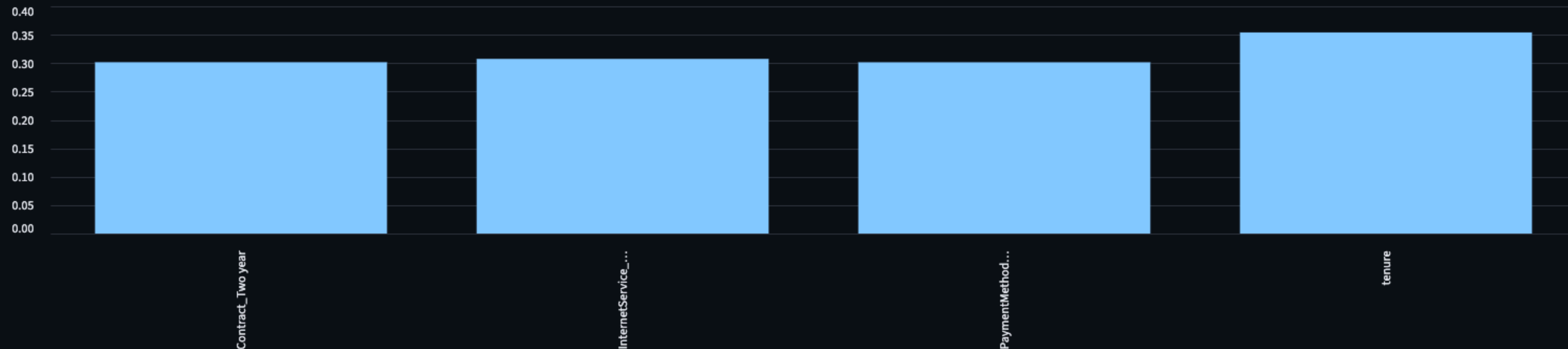
- Les clients récents (tenure faible) churnent davantage
- Les frais mensuels élevés augmentent le risque de churn
- Contrats longs (1–2 ans) réduisent fortement le churn
- Paiement par Electronic check → plus de churn
- Services OnlineSecurity & TechSupport → réduisent le churn

Piechart de la repartition du churn



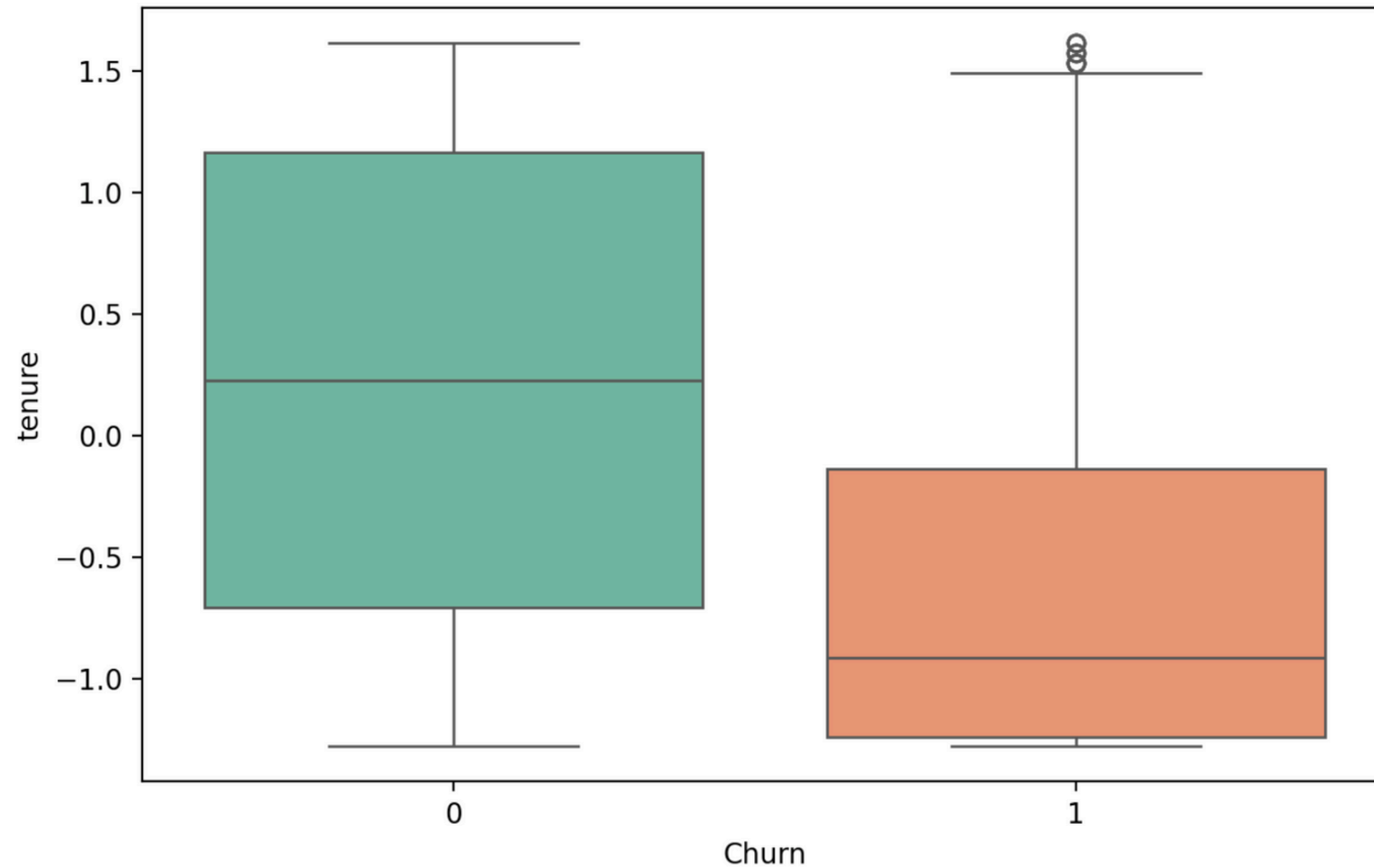
73.5 pourcent des clients sont restés (Churn = 0) 26.6 pourcent des clients ont churné (Churn = 1)=> soit 1 client sur 3,8

Top 4 variables les plus corrélées au churn



- Contrat Two Years
- InternetService_fiber_Optic
- PaymentMethod_Electronic_check
- tenure (plus élevé)

Ancienneté (tenure) selon le Churn



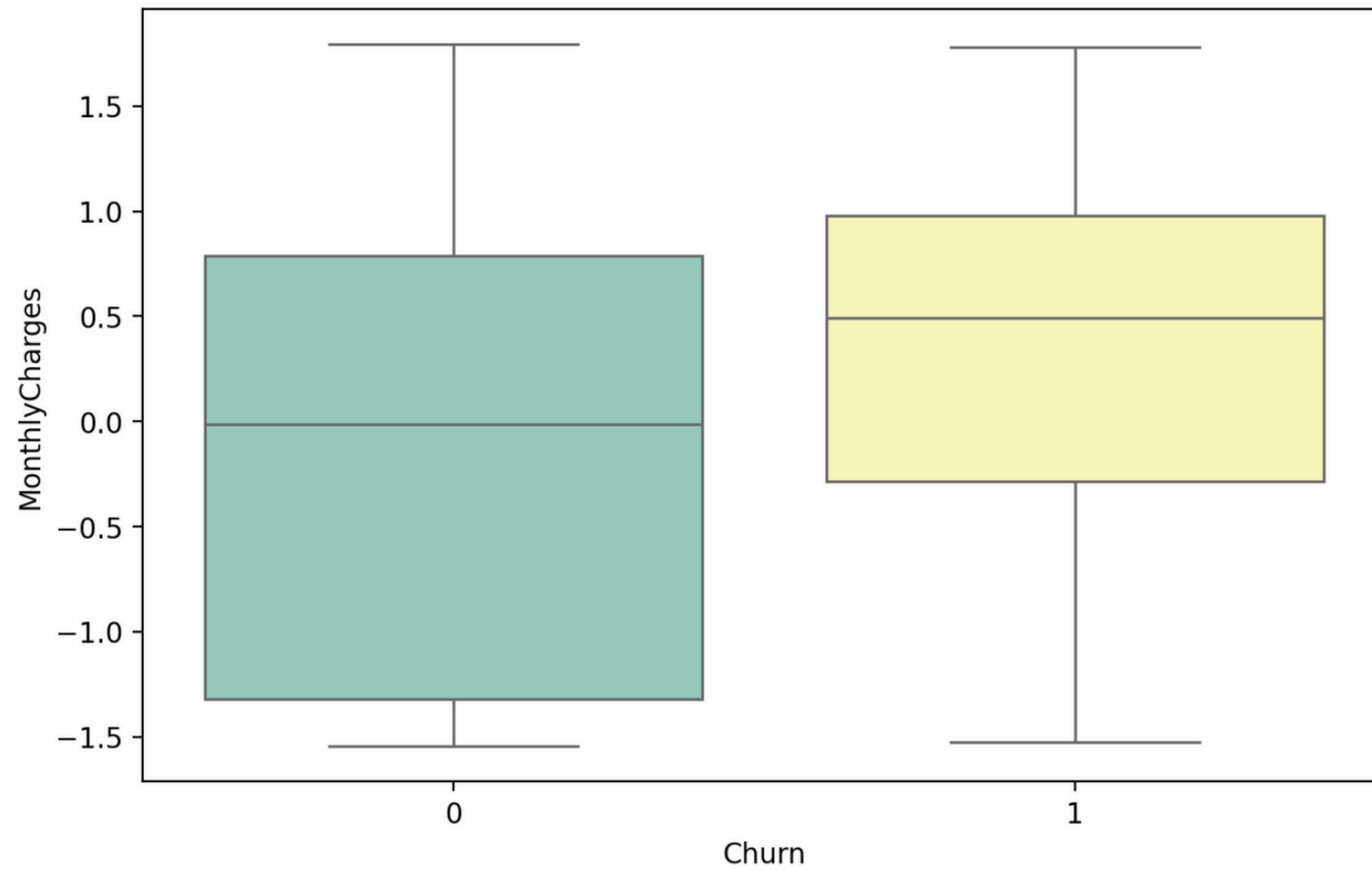
Ancienneté moyenne : Churn

No 0.213019 > 0

Yes -0.588451 < 0

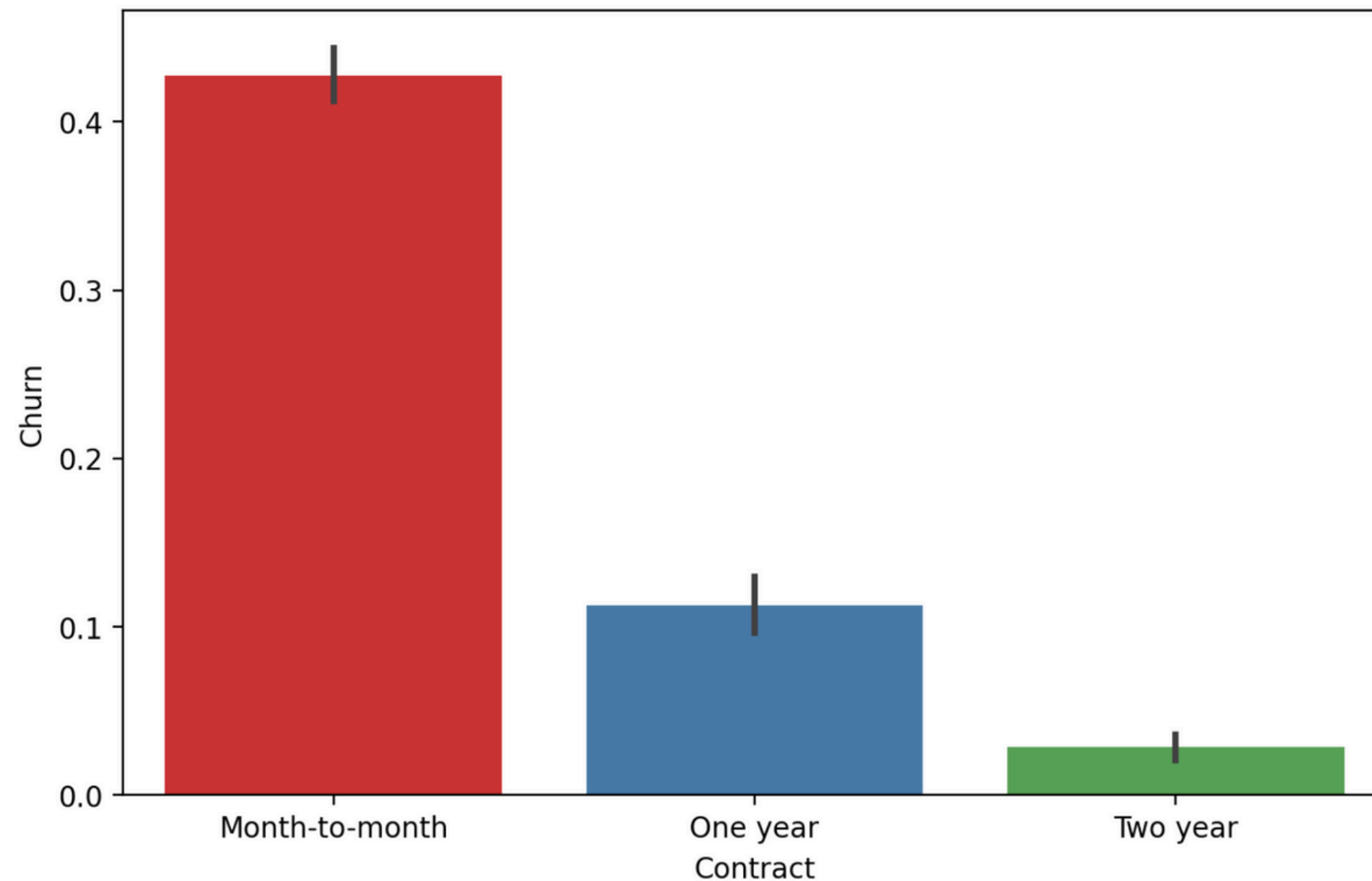
Les clients récents (tenure faible) churnent beaucoup plus que les anciens. Ils quittent l'entreprise tôt dans le cycle client.

Frais mensuels (MonthlyCharges) selon le churn



MonthlyCharges moyen : Churn
No $-0.116036 < 0$
Yes $0.320542 > 0$
Les clients avec des frais mensuels élevés
sont plus enclins à churner.

Type de contrat et churn



Taux de churn selon Contract_One year :

Contract_One year

0 30.629496

1 11.277174

Taux de churn selon Contract_Two year :

Contract_Two year

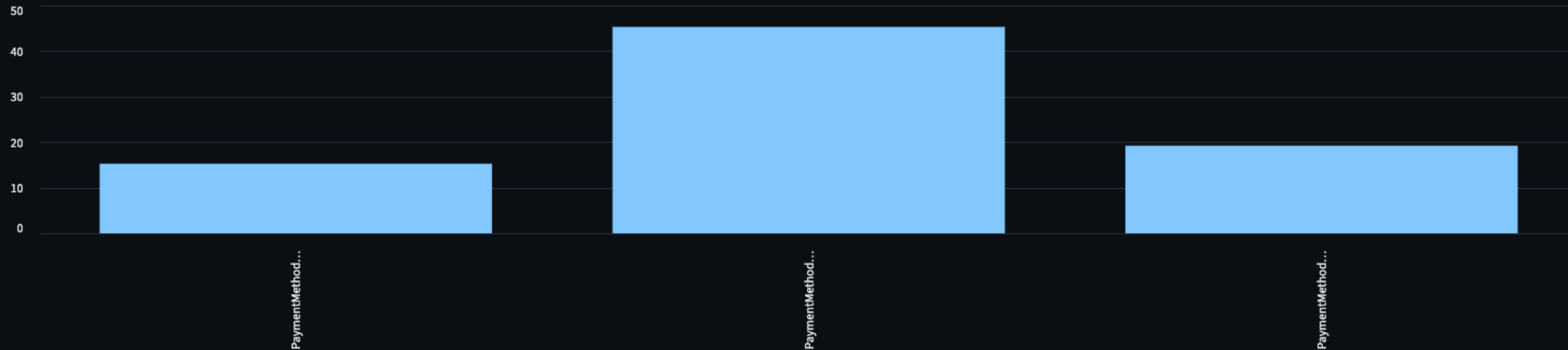
0 34.056480

1 2.848665

- Les clients en contrat 1 an → leur taux de churn chute à 11%.
- Les clients en contrat 2 ans → ils sont presque captifs, avec un churn ultra faible à 2,8%.
- Les clients en month-to-month sont ceux qui churnent le plus ≈ 40%,

Les contrats longs (1-2 ans) réduisent fortement le churn comparé aux contrats mensuels.

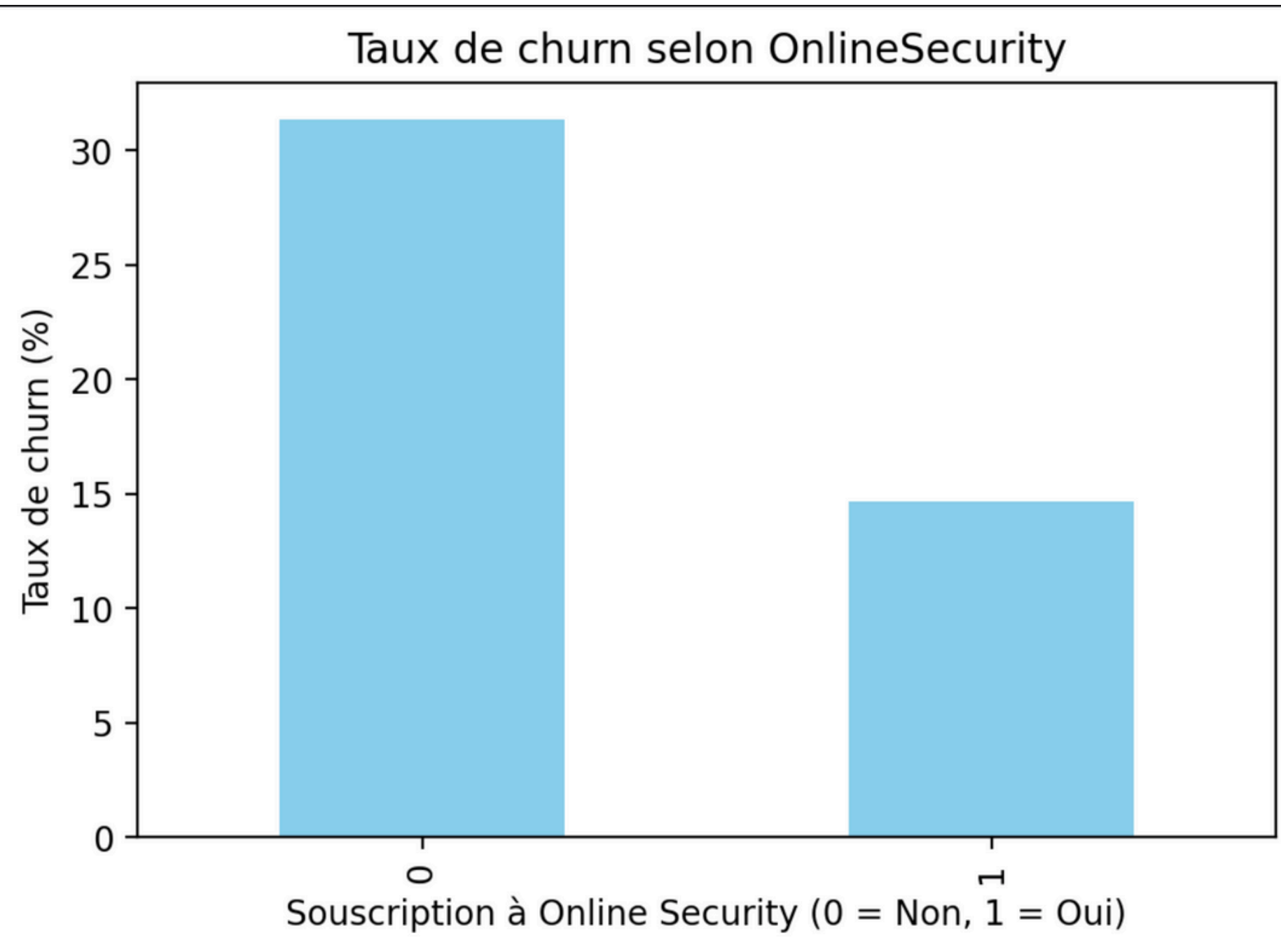
Méthode de paiement et churn ↗



- Payment method electronic check (CHURN élevé) => 45,3% => cheque électronique => méthode à risque
- Payment method credit card (CHURN faible) => 15,3% => carte bancaire => paiement automatisé = meilleure fidélité = moins de churn
- Payment method electronic check (CHURN moyen) => 19,2% => chèque postal => méthode intermédiaire

Les clients payant par Electronic check sont les plus à risque de churn.

OnlineSecurity et Churn



- Les clients qui n'ont pas activé OnlineSecurity sont beaucoup plus susceptibles de churner.
- Les clients protégés par OnlineSecurity restent plus fidèles.

Churn avec OnlineSecurity=1 : 14.6% vs sans OnlineSecurity=0 : 31.4%.

Résultats des modèles de Machine Learning

Les Modèles

	Modèle	Principe
0	Régression Logistique	Modèle linéaire qui estime la probabilité de churn en fonction des variables.
1	Arbre de Décision	Applique des règles successives (si/alors) pour classer un client.
2	Random Forest	Ensemble d'arbres de décision, combine leurs prédictions pour être plus robuste.

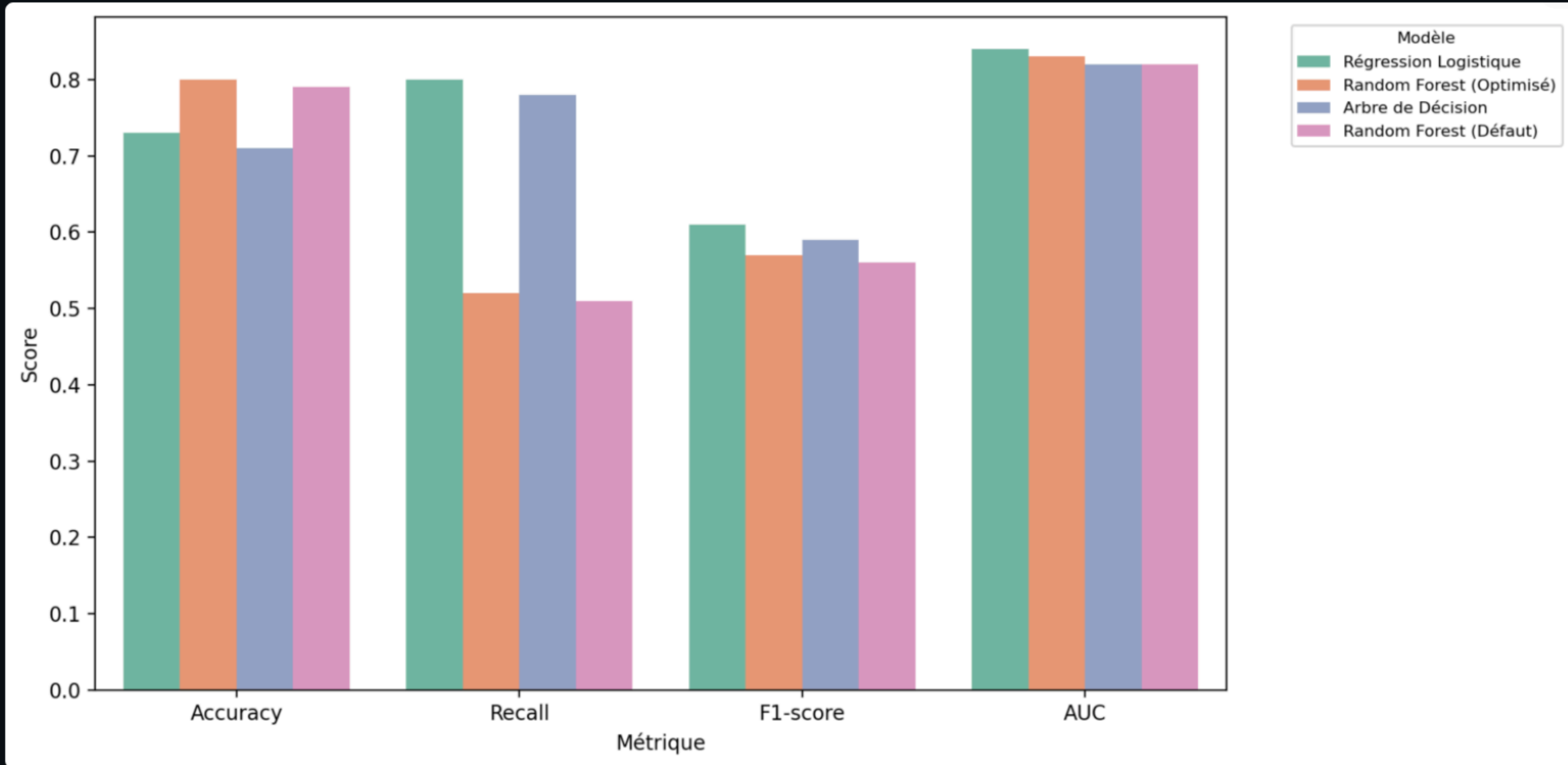
Métriques de performance

	Métrique	Définition
0	Accuracy	Proportion de prédictions correctes (globalement).
1	Recall	Capacité à détecter les churners (clients partis).
2	F1-score	Moyenne harmonique entre précision et rappel (équilibre).
3	AUC	Mesure globale de séparation entre churn et non-churn (0.5 = hasard, 1 = parfait).

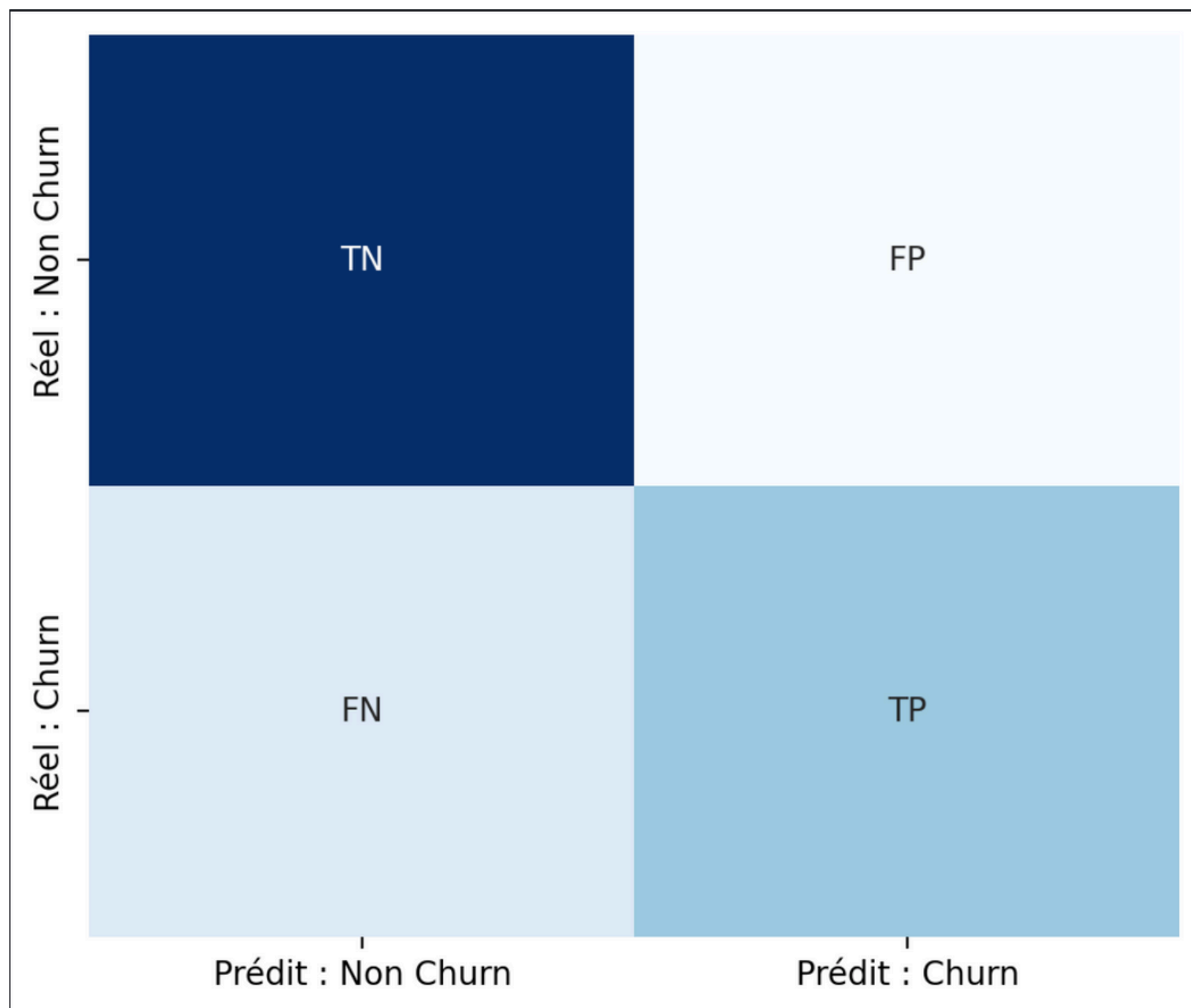
Comparaison des modèles

	Modèle	Accuracy	Recall	F1-score	AUC
0	Régression Logistique	0.73	0.8	0.61	0.84
1	Random Forest (Optimisé)	0.8	0.52	0.57	0.83
2	Arbre de Décision	0.71	0.78	0.59	0.82
3	Random Forest (Défaut)	0.79	0.51	0.56	0.82

Comparaison visuelle des performances



Matrices de confusion

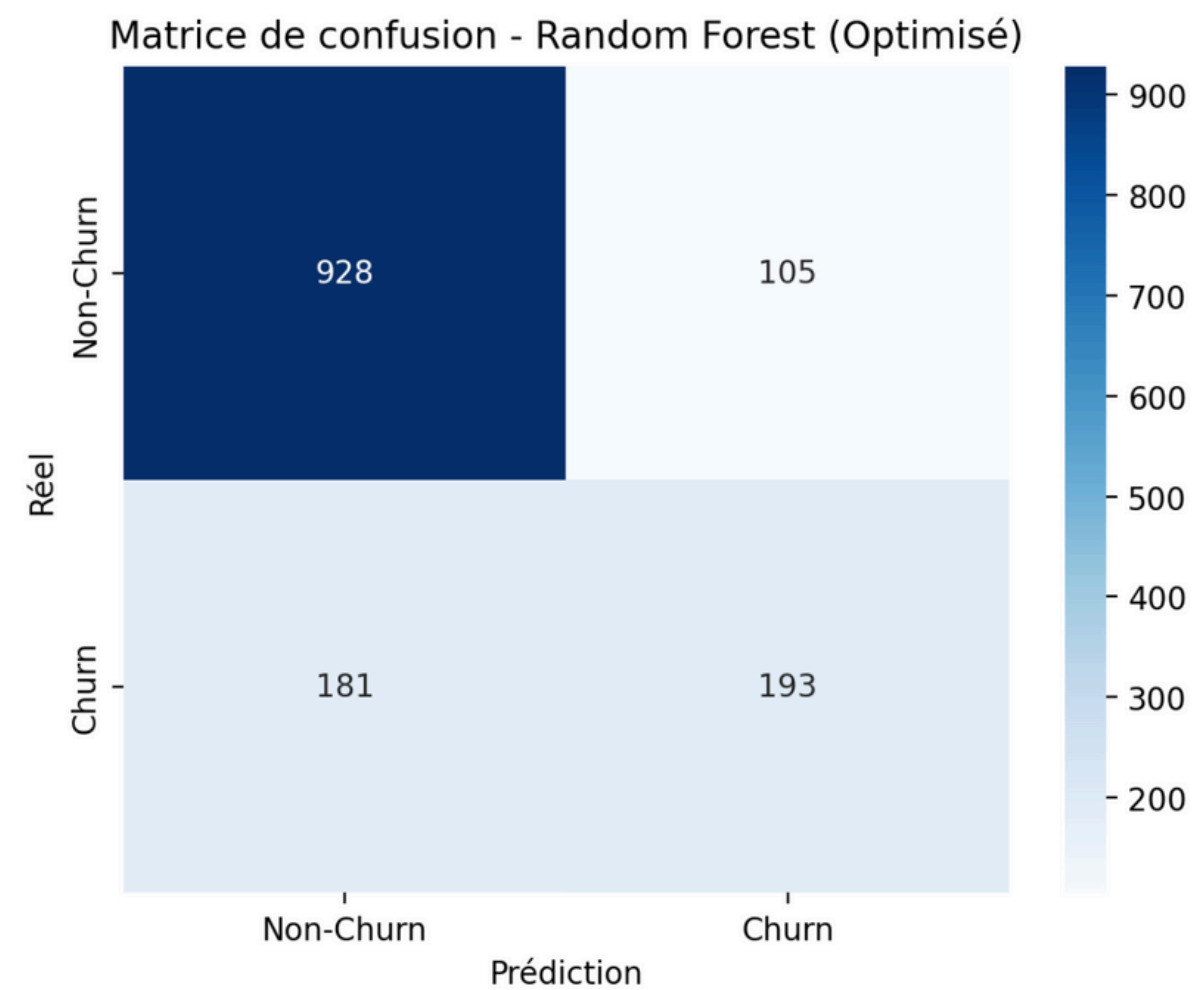
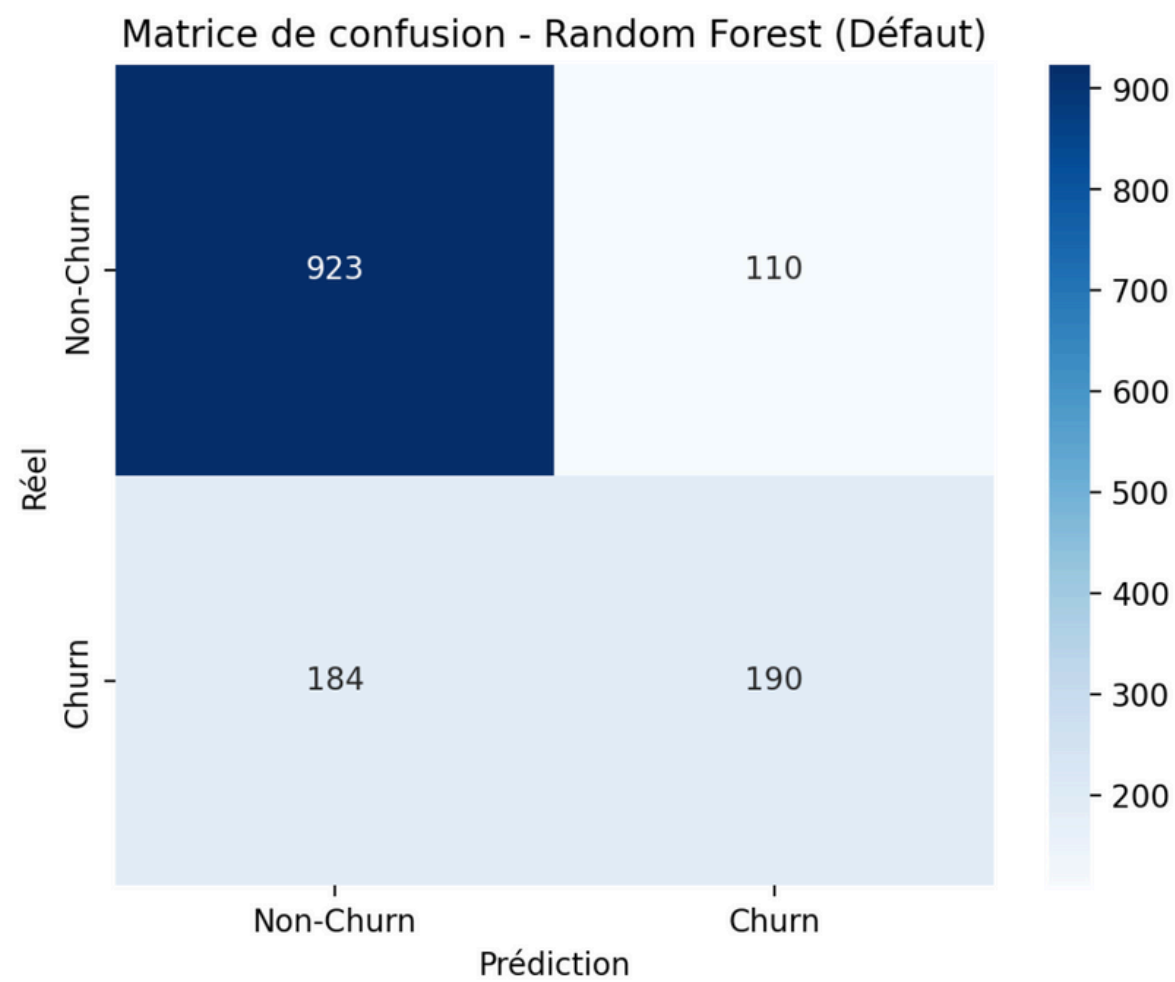
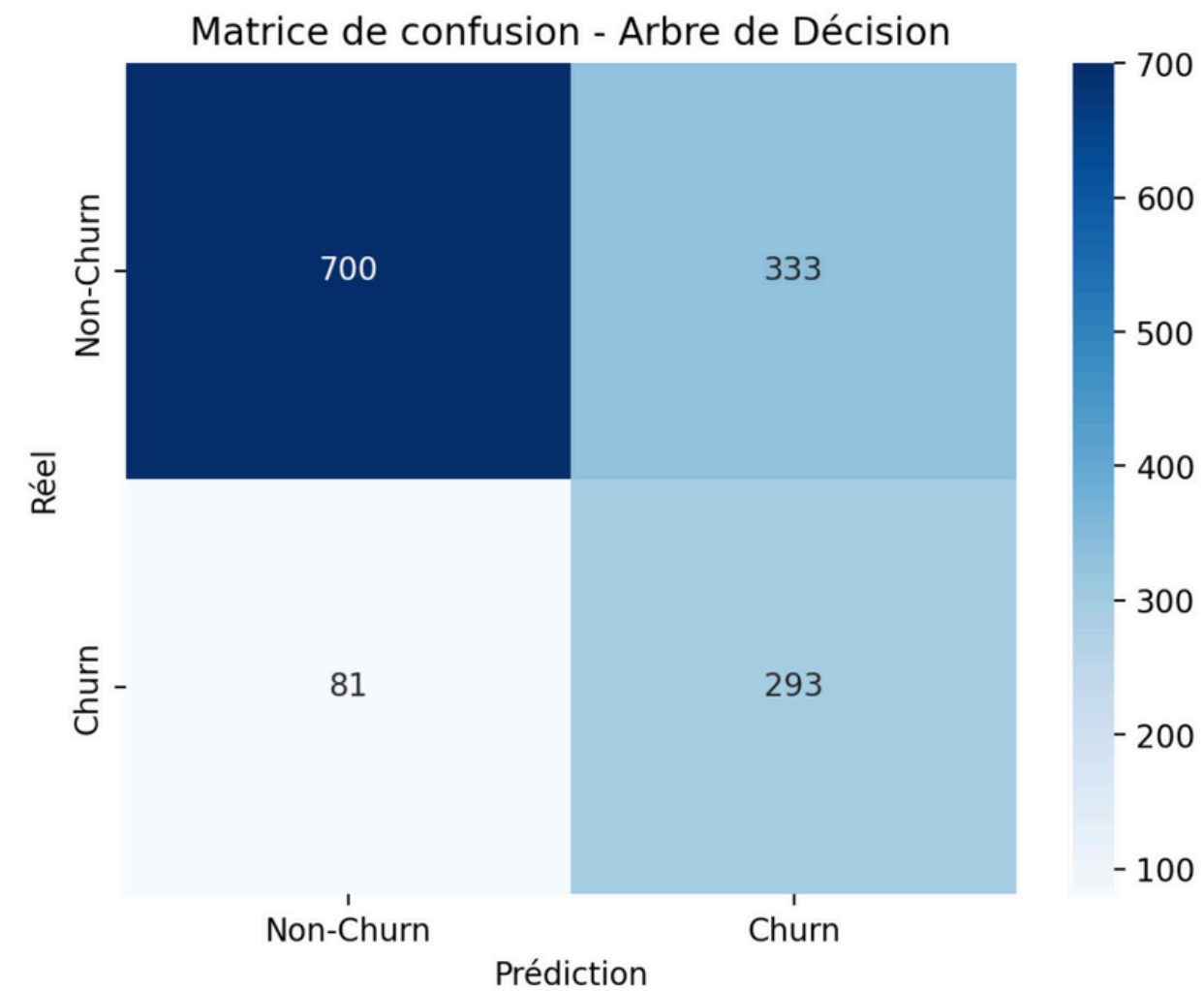
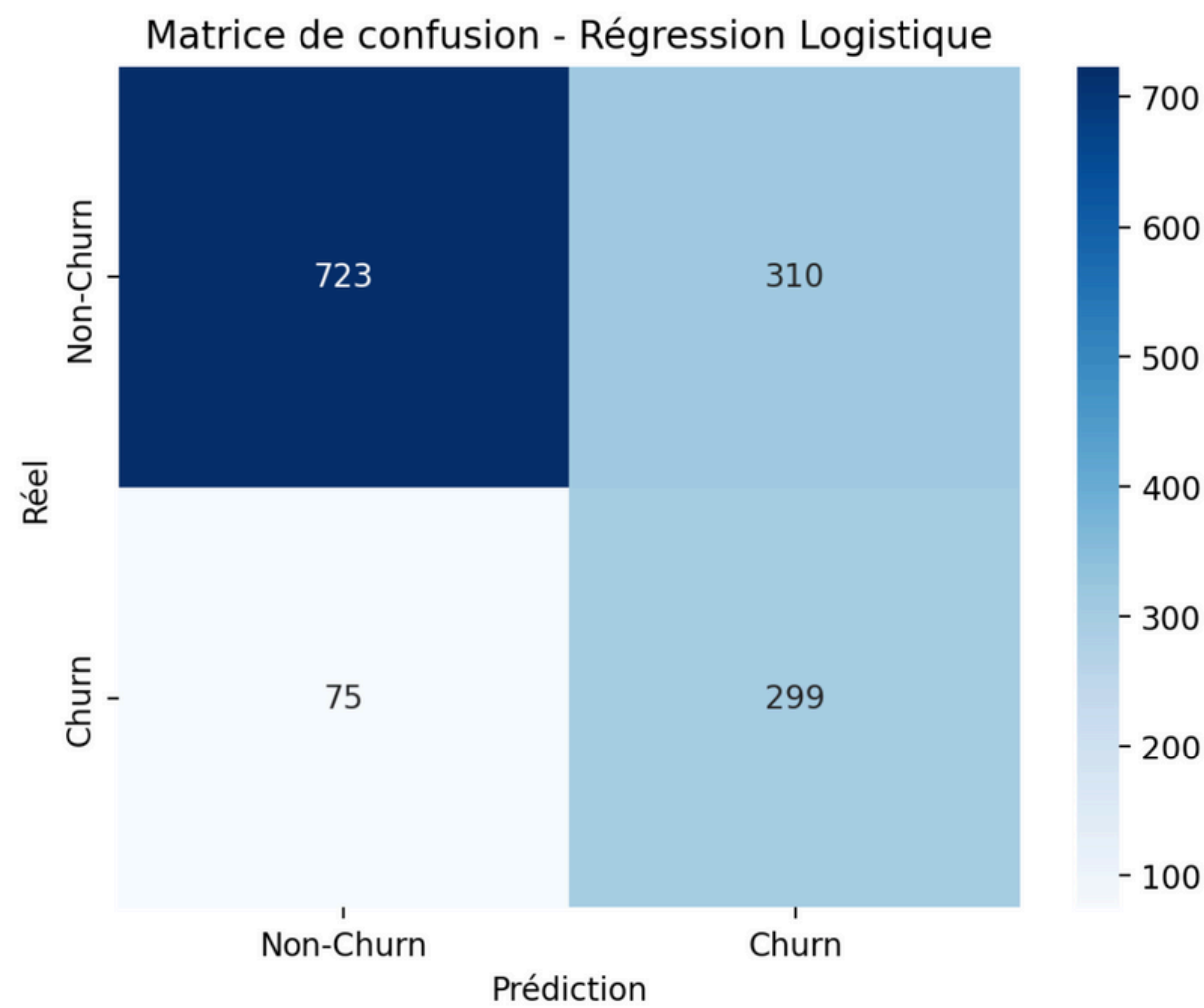


La matrice de confusion permet de comprendre comment un modèle se trompe :

- Trop de FP → trop d’alertes inutiles → risque d’agacer les clients.
- Trop de FN → beaucoup de churners ratés → perte de revenus.

on privilégie la réduction des FN (augmenter le Recall) car l’objectif est de détecter un maximum de churners.

	Type	Définition
0	Vrais Négatifs (TN)	Clients non churn correctement prédits comme non churn.
1	Faux Positifs (FP)	Clients non churn prédits à tort comme churn (fausse alerte).
2	Faux Négatifs (FN)	Clients churn prédits à tort comme non churn (churn ratés).
3	Vrais Positifs (TP)	Clients churn correctement prédits comme churn.



Conclusion et modèle retenu

- Régression Logistique : meilleur Recall (0.80) et meilleur AUC (0.84) → excellent pour détecter les churners
- Random Forest optimisé : meilleure Accuracy (0.80), mais Recall limité (0.52) → plus précis mais moins sensible
- Arbre de Décision : rappel élevé (0.78) mais performance globale plus faible
- Random Forest défaut : similaire à la version optimisée mais un peu moins stable

Modèle retenu : Régression Logistique, car l'objectif principal est de maximiser la détection des churners.